DOCUMENT RESUME

ED 041 931 24 TM 000 009

AUTHOR	Livingston, Samuel A.
TITLE	Some Observations on the Estimation of True Scores.
INSTITUTION	Johns Hopkins Univ., Baltimore, Md. Center for the Study of Social Organization of Schools.
SPONS AGENCY	Research and Development Centers Branch, DEL.
REPORT NO	JHU-69
BUREAU NO	BR-6-1610
PUB DATE	Jun 70
GRANT	OEG-2-7-061610-0207
NOTE	15p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.85
DESCRIPTORS	Analysis of Variance, Measurement, *Raw Scores, *Scores, *Statistical Analysis, Test Interpretation, *Test Reliability, *True Scores

ABSTRACT

The procedure of estimating true scores by means of a transformation of the obtained score based on the reliability coefficient is compared with the use of the obtained score without transformation. Using the mean squared error as a criterion, the transformed score is a better estimate for most examinees but poorer for those whose true scores lie furthest from the mean. An example is provided. (Author)

ED041931

TM
BR 6-1610
PA 24

Tm000009

THE JOHNS HOPKINS UNIVERSITY

REPORT No. 69

THE CENTER FOR THE STUDY OF SOCIAL ORGANIZATION OF SCHOOLS

SOME OBSERVATIONS ON THE ESTIMATION OF TRUE SCORES

BY

SAMUEL A. LIVINGSTON

JUNE, 1970

STAFF

John L. Holland, Director

James M. McPartland, Assistant Director

Virginia Bailey
Thelma Baldwin
Zahava D. Blum
Judith P. Clark
Karen C. Cohen
James S. Coleman
Robert L. Crain
David DeVries
Keith Edwards
Doris R. Entwisle
James Fennessey
Catherine J. Garvey
Ellen Greenberger
John T. Guthrie
Rubie Harris
Edward J. Harsch
Robert T. Hogan
Marian Hoover
Thomas Houston
Michael Inbar
Nancy L. Karweit
Judith Kennedy
Steven Kidder
Hao-Mei Kuo
Samuel Livingston
Edward L. McDill
Rebecca J. Muraro
Jeanne O'Connor
Martha O. Roseman
Peter H. Rossi
Joan Sauer
Leslie Schnuelle
Christine Schulze
Aage B. Sørensen
Annemette Sørensen
Julian C. Stanley
Clarice S. Stoll
Mary Viernstein
Murray A. Webster
Barbara J. Williams
Phyllis K. Wilson

ED041931

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Some Observations on the Estimation of True Scores

Grant No. -- OEG-2-7-061610-0207

Samuel A. Livingston

June, 1970

Prepared under the auspices of the Center for the Study of Social Organization of Schools, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this paper do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

The Johns Hopkins University

Baltimore, Maryland

ACKNOWLEDGEMENTS

I thank Gary A. Chase and Julian C. Stanley for their assistance in developing the ideas presented here.

Some Observations on the Estimation of True Scores

Given the obtained score of one person on one form of a test, what is the best estimate of his true score? Obtained scores in a group of examinees will tend to covary positively with the errors of measurement that they contain. To compensate for this fact, we usually estimate the true score by linear regression. Expressing both the obtained score and the true score as deviations from the mean of the group of persons tested (the mean of their obtained scores will be equal to the mean of their true scores), the estimate is $\rho^2_{tx}x_{pf}$.* In this notation x_{pf} is the obtained (deviation) score of person $\underline{p}$ on form $\underline{f}$, and ρ^2_{tx} is the squared correlation (over the whole group of persons) of true scores with obtained scores, which is equal to the reliability coefficient of the test. (For a more detailed exposition of this method of estimating true scores, see Stanley, 1970.)

The problem with this estimate is that its expectation over forms for a single person is not equal to t_p , the person's true (deviation) score. Therefore one cannot construct a confidence interval for t_p which will center around $\rho^2_{tx}x_{pf}$. In fact, the only estimate of t_p

---

*Lord (1959) points out that this estimate is actually only an approximation; the correct regression coefficient for the least-squares linear-regression estimate is considerably more complicated.

which does have an expectation (over forms) equal to t_p is x_{pf}, the person's obtained (deviation) score.

This is the dilemma: When we make a point estimate of t_p, we use $\rho^2_{tx}x_{pf}$ as our estimate, but when we make an interval estimate of t_p, we center the interval around x_{pf}.

It is possible to center a confidence interval around $\rho^2_{tx}x_{pf}$, but this interval will <u>not</u> be a confidence interval for the person's true score. Instead, it will be a confidence interval for the true-score component of his estimated-true-score; that is, not for t_p but for $\rho^2_{tx}t_p$.

Which is the better estimate of t_p, x_{pf} or $\rho^2_{tx}x_{pf}$? We can compare them by using as a criterion the mean squared error of estimate over forms (abbreviated MSE).

The MSE for x_{pf} is simply the variance (over forms) of errors of measurement for person p:

$$(1)\quad \text{MSE}\,(x_{pf}) = E_f(x_{pf} - t_p)^2 = E_f(e_{pf})^2 = \sigma^2(e_{p*}) ,$$

where the asterisk as the second subscript for e indicates that the variance is computed over forms.

The expression for the MSE for $\rho^2_{tx}x_{pf}$ is somewhat more complicated. It can be derived as follows:

(2) MSE $(\rho^2_{tx}x_{pf}) = E_f(\rho^2_{tx}x_{pf} - t_p)^2$

$= E_f(\rho^4_{tx}x^2_{pf} + t^2_p - 2\rho^2_{tx}t_p x_{pf})$

$= \rho^4_{tx}E_f(x^2_{pf}) + t^2_p - 2\rho^2_{tx}t_p E_f(x_{pf})$.

We know that $E_f(x_{pf})$ is simply t_p . What is $E_f(x^2_{pf})$? It is

(3) $E_f(t_p + e_{pf})^2 = E_f(t^2_p + e^2_{pf} + 2t_p e_{pf})$

$= t^2_p + E_f(e^2_{pf}) + 2t_p E_f(e_{pf}) = t^2_p + \sigma^2(e_{p*})$,

since $E_f(e_{pf}) = 0$.

From (2) and (3), we then have

(4) MSE $(\rho^2_{tx}x_{pf}) = \rho^4_{tx}[t^2_p + \sigma^2(e_{p*})] + t^2_p - 2\rho^2_{tx}t^2_p$

$= \rho^4_{tx}\sigma^2(e_{p*}) + t^2_p + \rho^4_{tx}t^2_p - 2\rho^2_{tx}t^2_p$

$= \rho^4_{tx}\sigma^2(e_{p*}) + (1 - \rho^2_{tx})^2 t^2_p$.

The expectation of this quantity over persons (remember that x_{pf} and t_p are defined here as deviation scores) is

(5) $E_p[\rho^4_{tx}\sigma^2(e_{p*}) + (1 - \rho^2_{tx})^2 t^2_p]$

$= \rho^4_{tx}E_p\sigma^2(e_{p*}) + (1 - \rho^2_{tx})^2 E_p(t^2_p)$

$$= \rho_{tx}^4 \sigma_e^2 + (1 - \rho_{tx}^2)^2 \sigma_t^2$$

$$= \rho_{tx}^4 \sigma_e^2 + (1 - \rho_{tx}^2)^2 \rho_{tx}^2 \sigma_x^2$$

$$= \rho_{tx}^4 \sigma_e^2 + \rho_{tx}^2 (1 - \rho_{tx}^2) \sigma_x (1 - \rho_{tx}^2)$$

$$= \rho_{tx}^4 \sigma_e^2 + \rho_{tx}^2 (1 - \rho_{tx}^2) \sigma_e^2$$

$$= \rho_{tx}^2 \sigma_e^2 [\rho_{tx}^2 + (1 - \rho_{tx}^2)]$$

$$= \rho_{tx}^2 \sigma_e^2 \ .$$

Since $\rho_{tx}^2 < 1$, the MSE for $\rho_{tx}^2 x_{pf}$ over all persons in the group is less than that for x_{pf}. Is there any identifiable subgroup of persons for whom x_{pf} is a better estimator of t_p than is $\rho_{tx}^2 x_{pf}$? If so, these will be the persons for whom

MSE (x_{pf}) < MSE $(\rho_{tx}^2 x_{pf})$; that is, for whom

$$\sigma^2(e_{p*}) < \rho_{tx}^4 \sigma^2(e_{p*}) + (1 - \rho_{tx}^2)^2 t_p^2 \ ,$$

where $\sigma^2(e_{p*})$ is the variance of errors of measurement over forms for person p. If we are willing to assume that any person's error variance (over forms) is the same as the variance of errors of measurement over persons on a given form, we can substitute $\sigma^2(e_{*f})$ for $\sigma^2(e_{p*})$, where $\sigma^2(e_{*f})$ is the variance of errors of measurement over persons

on form $\underline{f}$. Thus

$$\sigma^2(e_{*f}) < \rho_{tx}^4\sigma^2(e_{*f}) + (1 - \rho_{tx}^2)^2 t_p^2 .$$

Using σ_e^2 to represent $\sigma^2(e_{*f})$, we have

$\sigma_e^2(1 - \rho_{tx}^4) < (1 - \rho_{tx}^2)^2 t_p^2$. Therefore, $\sigma_e^2(1 + \rho_{tx}^2) < (1 - \rho_{tx}^2)t_p^2$.

$[\sigma_x^2(1 - \rho_{tx}^2)](1 + \rho_{tx}^2) < (1 - \rho_{tx}^2)t_p^2$.

Then, if $\rho_{tx}^2 \neq 0$, $\sigma_x^2(1 + \rho_{tx}^2) < t_p^2$.

$$(6) \quad \frac{\sigma_x^2}{\sigma_t^2}(1 + \rho_{tx}^2) < \frac{t_p^2}{\sigma_t^2} .$$

$$\frac{1}{\rho_{tx}^2}(1 + \rho_{tx}^2) < (\frac{t_p}{\sigma_t})^2 .$$

Therefore, for any give value of ρ_{tx}^2 , x_{pf} will be a better estimator of t_p than will $\rho_{tx}^2 x_{pf}$ for those persons whose true scores lie furthest from the mean. Notice that as ρ_{tx}^2 approaches zero, the expression on the left becomes very large. This means that the less reliable the test, the smaller the proportion of examinees for whom x_{pf} is the better estimator. Also notice that since $\rho_{tx}^2 < 1$, the expression on the left in inequality (6) is always greater than 2. Therefore the "cutoff points" will always be greater

than $\sigma_t\sqrt{2}$ from the mean. Consequently, the proportion of examinees for whom x_{pf} is the better estimator will not exceed 16 per cent.

One important problem remains, however. We do not know the true score of any examinee--only his obtained score. How, then, can we estimate t_p/σ_t in order to know which estimate of t_p to use? Clearly if <u>both</u> estimates of t_p satisfy inequality (6), then we should use x_{pf}, while if <u>neither</u> estimate of t_p satisfies inequality (6), then we should use the transformation $\rho^2_{tx}x_{pf}$.

Since $\rho^2_{tx} \leq 1$, the use of the transformation will always produce a smaller value for the right side of inequality (6). Therefore, if the inequality holds when $\rho^2_{tx}x_{pf}$ is used as the estimate of t_p, it will hold when x_{pf} is used as the estimate. Also, if the inequality does not hold when x_{pf} is used as the estimate of t_p, it will not hold when $\rho^2_{tx}x_{pf}$ is used. We can estimate σ^2_t by $\rho^2_{tx}\sigma^2_x$. Therefore, we can construct two sets of cutoff points based on obtained scores. We should use x_{pf} as the estimate of t_p for any person for whom inequality (6) holds when $\rho^2_{tx}x_{pf}$ is substituted for t_p:

$$\frac{1}{\rho^2_{tx}}(1 + \rho^2_{tx}) < \frac{(\rho^2_{tx}x_{pf})^2}{\rho^2_{tx}\sigma^2_x};$$

$$\frac{1}{\rho^2_{tx}}(1 + \rho^2_{tx}) < \frac{\rho^2_{tx}x^2_{pf}}{\sigma^2_x};$$

$$\frac{\sigma_x^2}{\rho_{tx}^4}(1 + \rho_{tx}^2) < x_{pf}^2 ,$$

that is, for any person whose obtained score is more than

$$\frac{\sigma_x\sqrt{1 + \rho_{tx}^2}}{\rho_{tx}^2}$$ units from the mean.

We should use $\rho_{tx}^2 x_{pf}$ as the estimate of t_p for any person for whom inequality (6) does <u>not</u> hold when x_{pf} is substituted for t_p :

$$\frac{1}{\rho_{tx}^2}(1 + \rho_{tx}^2) \leq \frac{x_{pf}^2}{\rho_{tx}^2\sigma_x^2} ;$$

$$\sigma_x^2(1 + \rho_{tx}^2) \leq x_{pf}^2 ;$$

that is, for any person whose obtained score is not more than

$\sigma_x\sqrt{1 + \rho_{tx}^2}$ units from the mean.

Consider as an example a test with a reliability of .90. In this case the cutting points would lie at $\pm 1.38\sigma_x$ and $\pm 1.53\sigma_x$. (In a normal distribution these points would identify the middle 83.2% of the examinees and the highest 6.3% and lowest 6.3%.)

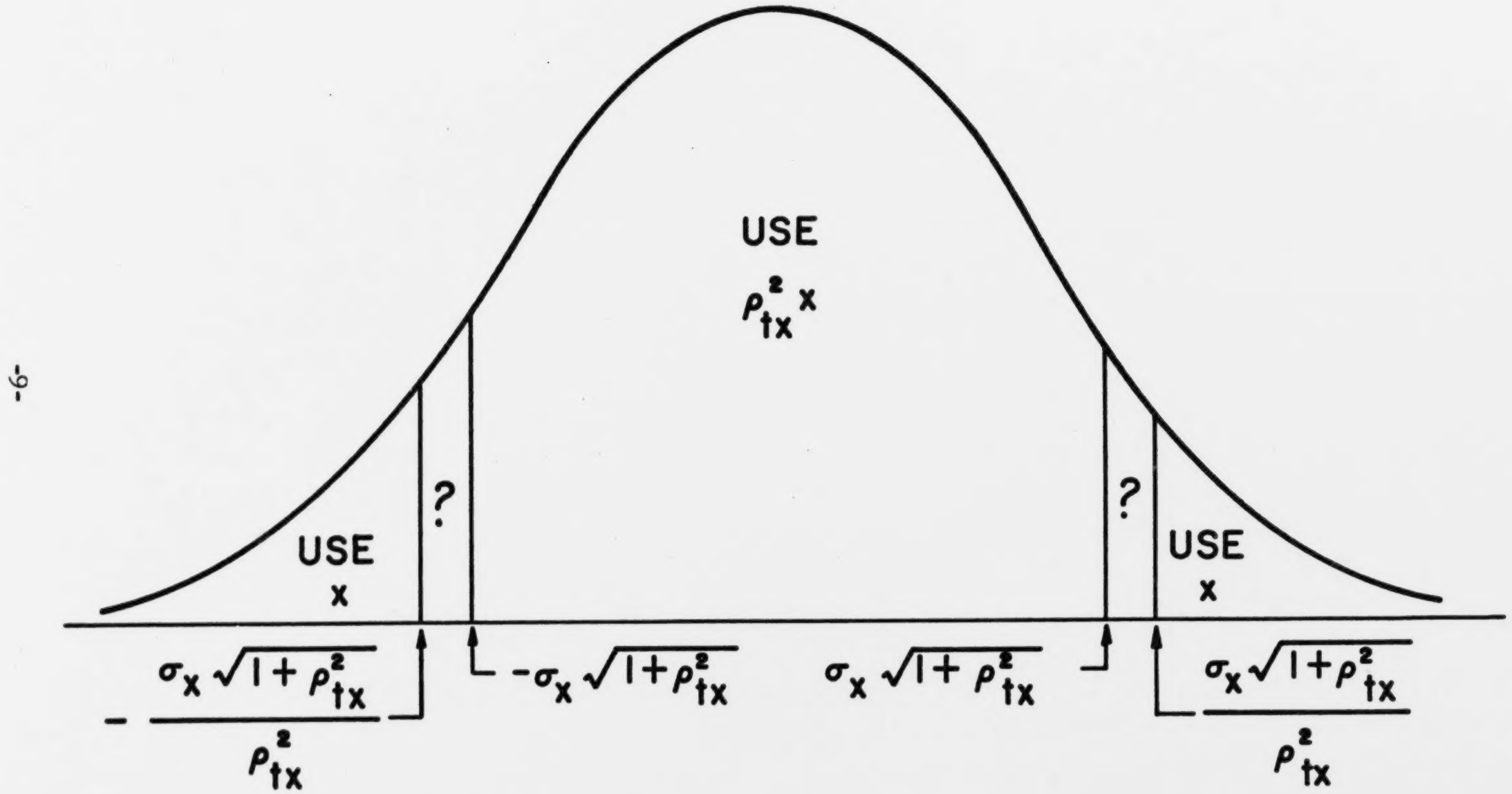
What about those examinees whose obtained scores lie between the two sets of cutting points? (In the example these would be 4.2% of all the examinees.) What is the best estimate of their true scores?

Unless we are willing to estimate their true scores by more complicated curvilinear-regression methods* (in which case the whole discussion presented here becomes unnecessary), we can say only that the best estimate of t_p lies between x_{pf} and $\rho^2_{tx}x_{pf}$ -- between the obtained score and its transformed value. Figure 1 illustrates this situation graphically.

---

*See Lord (1959)

Figure 1.



Which estimate of a person's true score should we use?

References

Lord, Frederic M. Statistical inferences about true scores. Psychometrika, 1959, 24, 1-17.

Stanley, Julian C. Reliability. In Thorndike, R. L. (Ed.) Educational Measurement. Washington: American Council on Education, 1970.

This Report may now be obtained through the ERIC system at 5 cents per page, plus state sales tax and 50 cents handling charges. You should mail these requests with the ERIC number to:

ERIC Document Reproduction Service
National Cash Register Company
4936 Fairmont Avenue
Bethesda, Maryland 20014